

# **Final Technical Report**

**Prepared for the U.S. Geological Survey  
National Geological and Geophysical Data Preservation Program  
Award No. G14AP00122**

**09/01/2014 – 09/01/2015**

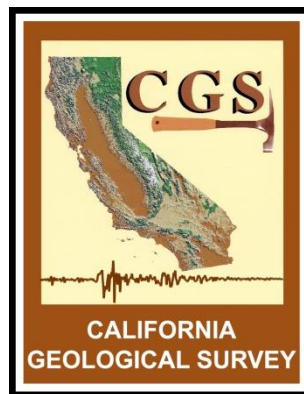
## **PRESERVATION AND DIGITAL CONVERSION OF GEOTECHNICAL REPORT AND BOREHOLE DATA CALIFORNIA GEOLOGICAL SURVEY, SACRAMENTO**

**Submitted October 30, 2015**

**By**

**Michael Silva  
Terilee McGuire  
California Geological Survey  
801 K Street, MS 12-31  
Sacramento, CA 95814  
Phone: (916) 324-0768  
FAX: (915) 445-3334  
E-mail: [Michael.Silva@conservation.ca.gov](mailto:Michael.Silva@conservation.ca.gov)**

**Department of Conservation  
California Geological Survey**



## ABSTRACT

Funds provided through the U.S. Geological Survey – National Geological and Geophysical Data Preservation Program (NGGDPP) Award No. G14AP00122, enabled the California Geological Survey (CGS) to support an ongoing **digital infrastructure project** focusing on geotechnical borehole data collected for CGS seismic hazard zonation. The goal of the project is to digitize all collected geotechnical reports used in zonation, and make the data available to the public through a web service.

Since 1996, CGS' Seismic Hazards Mapping Program has collected thousands of engineering geologic and geotechnical engineering reports throughout the state and has used the borehole and lab test data within these reports to produce regulatory Zones of Required Investigation. Data from these collected reports were entered into various databases and analyzed to identify areas susceptible to soil liquefaction or earthquake-induced landslides.

Over the last two years a majority of the original paper copies of the collected geotechnical reports have been scanned to PDF format. In addition, a new data model has been developed for the geotechnical data. CGS currently has borehole and laboratory geotechnical data in two or more databases, and scanned geotechnical reports are not linked to the borehole databases.

At present, only borehole data used for liquefaction zoning is available for public and governmental access online and the borehole data used for landslide hazard zone mapping is not available. The web site and the underlying technology are outdated and the data do not contain links to the scanned reports. As part of the NGGDPP FY-2014 **digital infrastructure** grant, CGS was able to 1) develop workflows to translate existing geotechnical data into the new data model, 2) generate new report level information for all geotechnical borings, and 3) link the report level data to the scanned PDFs and borings.

CGS selected a dataset limited to existing borehole data and scanned reports in Alameda County as the pilot project area. Approximately 1500 new report location records were generated for the landslide and liquefaction boreholes. The existing geotechnical data were translated into the new data model and a small subset of these data (data for the Dublin quadrangle) was published as a feature-based web service in ArcGIS as an initial proof of concept.

## TABLE OF CONTENTS

ABSTRACT .....	1
TABLE OF CONTENTS.....	2
LIST OF FIGURES.....	3
DATA PRESERVATION FOR THE CALIFORNIA GEOLOGICAL SURVEY:	
Introduction .....	4
Project Description.....	6
Results .....	9
Future .....	10

## LIST OF FIGURES

Figure 1. Sample geotechnical report for liquefaction.....	5
Figure 2. Screenshot of existing CGS web page.....	6
Figure 3. QGIS screenshot showing report forms and linked boreholes .....	7
Figure 4. New geotechnical data model.....	8
Figure 5. Screenshot of proof of concept web service.....	9

## DATA PRESERVATION FOR THE CALIFORNIA GEOLOGICAL SURVEY

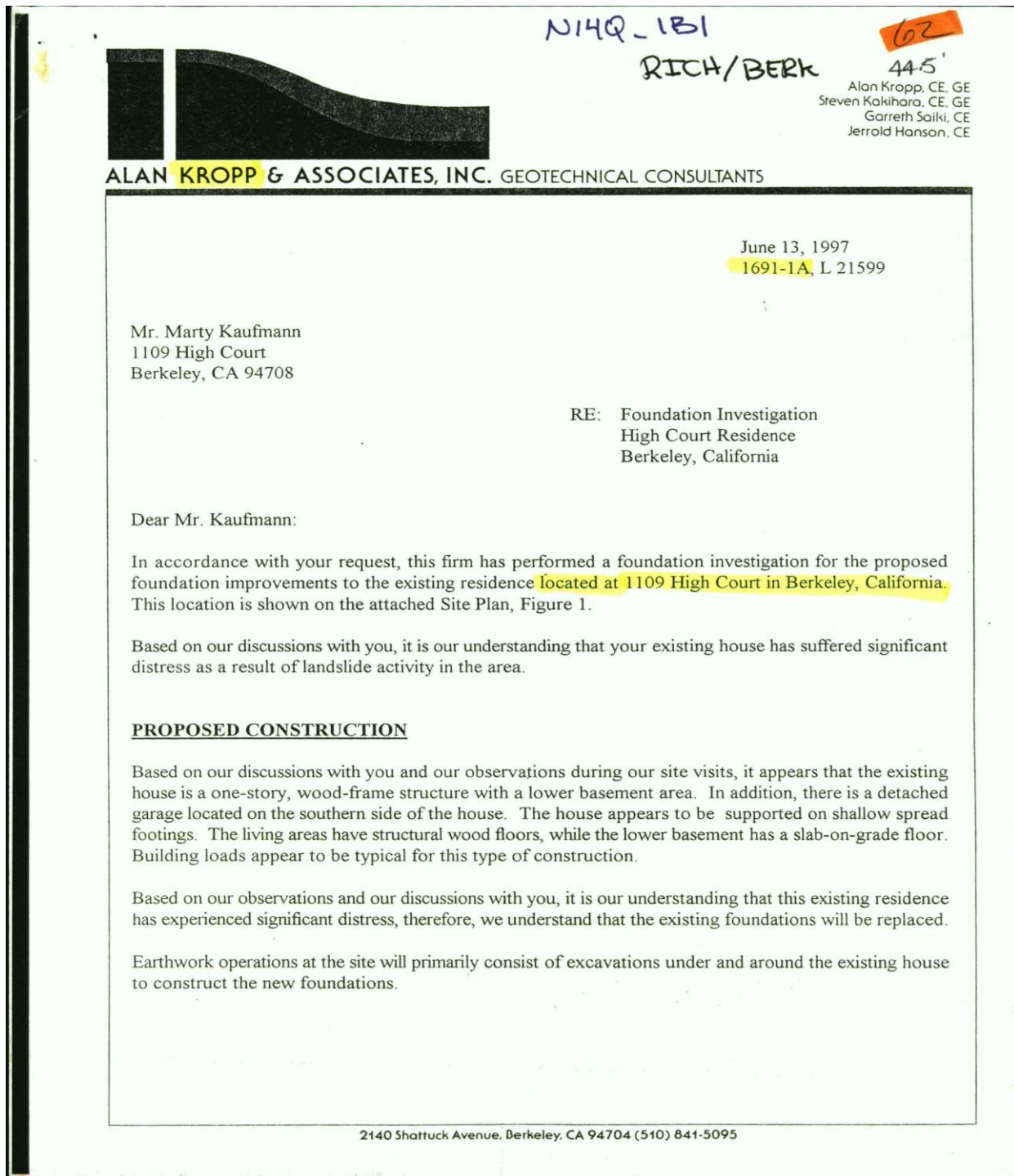
### Introduction:

This report covers the **digital infrastructure project** that was funded in part through the U.S. Geological Survey – National Geological and Geophysical Data Preservation Program (NGGDPP) Award No. G14AP00122.

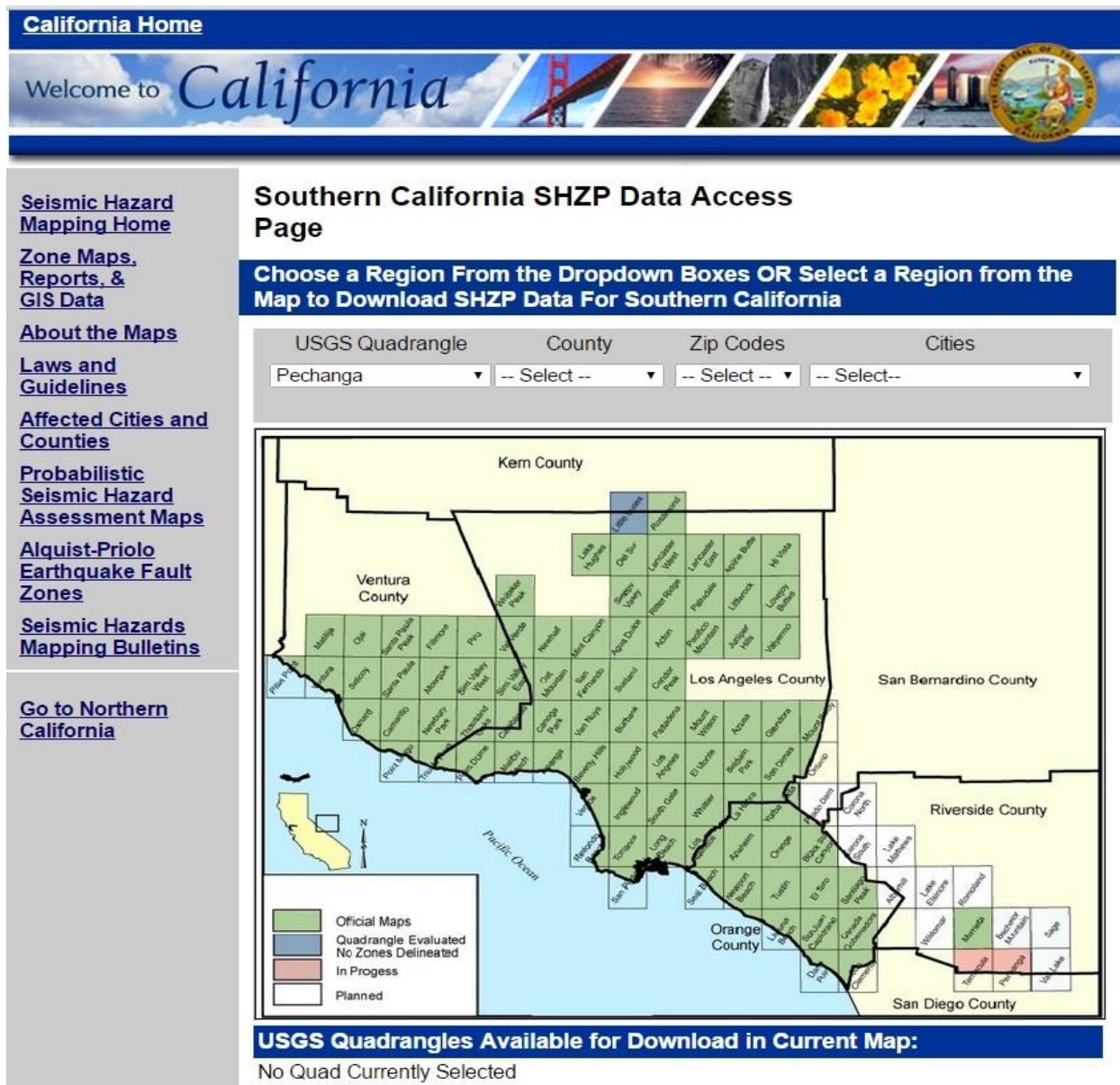
The Seismic Hazards Mapping Act was passed by the California Legislature in response to the 1989 Loma Prieta earthquake. This legislation requires the California Geological Survey to delineate Zones of Required Investigation, where the probability of ground failure due to soil liquefaction or landslides, is high. CGS determines the boundaries of these zones on the basis of data contained within engineering geologic and geotechnical engineering reports (Figure 1) prepared by consultants for development projects, submitted to local governmental agencies to obtain building permits. Since 1996, paper copies of reports have been collected by CGS staff and relevant field and laboratory test data were entered into databases and analyzed to delineate potentially hazardous areas.

Nearly 120 Seismic Hazard Zone maps have been prepared, covering 145 communities in eight counties, and more are currently in progress. Databases were developed in the mid-1990's to store the collected and processed geotechnical data, and more than 500,000 records are contained in these databases. A borehole database for liquefaction-related data was put online by CGS in 2003 (Figure 2), where users both within and outside of state government could access and download borehole data. However, the software running this web access is antiquated and no longer supported. An interim database is in development at this time. Borehole data associated with earthquake-induced landslides has not been migrated to an online release format and exists only in Microsoft Access.

In 2010, the Seismic Hazards Zoning Program started two major initiatives; first, a new data model was developed to house all the geologic/geotechnical data and, second, paper copies of all the collected report data started being scanned and saved as multipage PDFs. The new data model has a field for a hyperlink to the PDF so that users can ingest these data as either digital database files or view the PDF that they were derived from; the latter option being frequently requested by users outside of CGS. The development of the new unified database will provide an end point for data that currently exists in a number of physical and digital forms. This project uses the existing data for one county to develop translation workflows to migrate the legacy data into the final database and develop procedures and protocols for development of a new web service.



**Figure 1.** Cover sheet of typical geotechnical report collected for liquefaction hazard zoning.



**Figure 2.** Existing Seismic Hazard Zone borehole data web page. This server was put into service in 2003.

### Project Description:

This project began September 1, 2014 and the USGS contract expired on August 31, 2015. The project was limited to existing borehole data and scanned reports in Alameda County. This area was chosen for the completeness of data and the availability of previously scanned geotechnical reports.

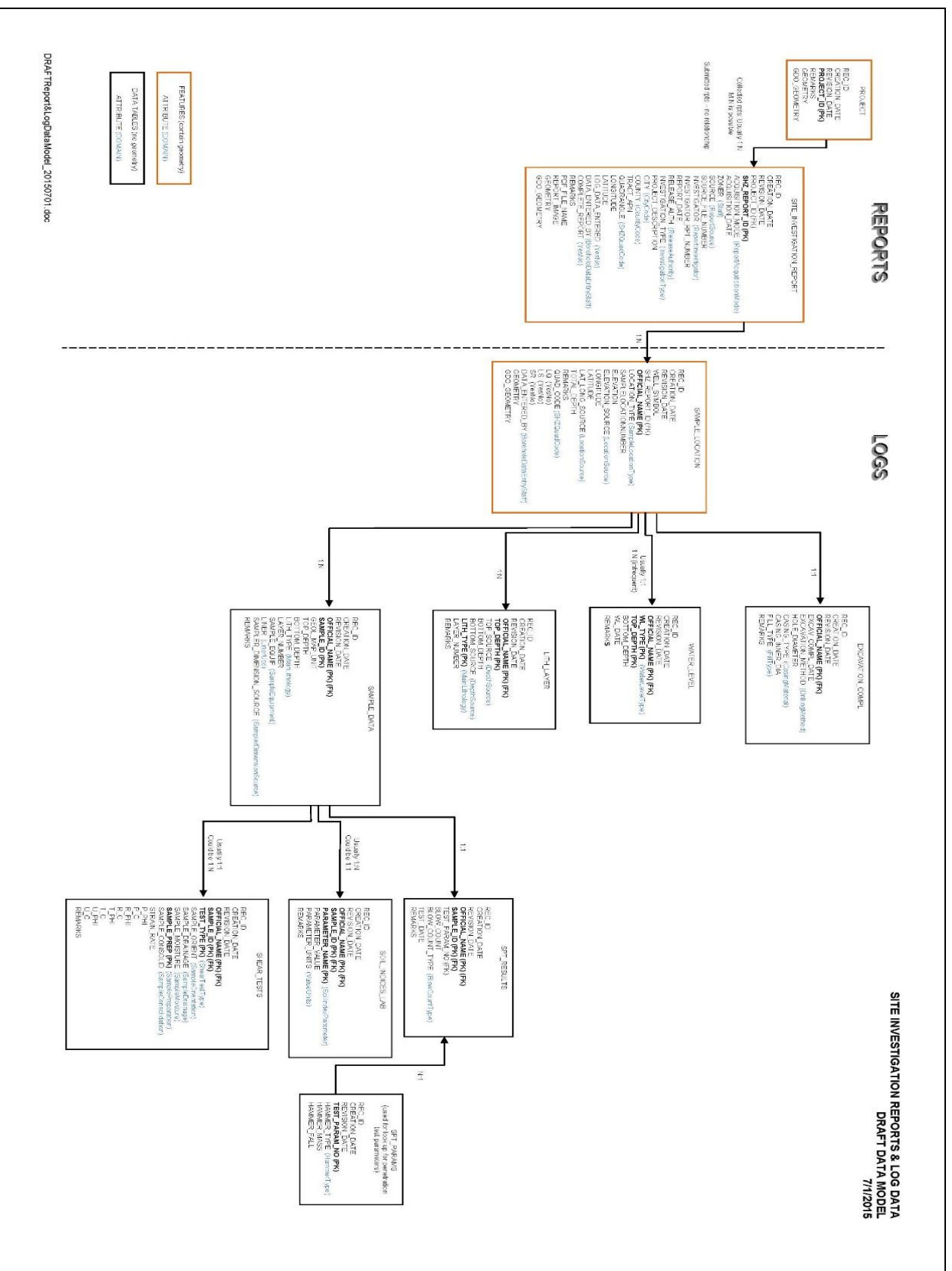


Initial project components consisted of over a thousand scanned geotechnical reports for both landslide and liquefaction hazard zoning in Adobe PDF format. Borehole data for landslide zoning existed in a multi-table Microsoft Access database. Borehole data for liquefaction zoning existed in a multi-table Oracle enterprise database. The new data model developed for CGS borehole data adds a new primary report table, which includes links to the borehole data tables and a scanned PDF geotechnical report located on CGS servers (Figures 3 & 4). The new data structure is designed for hosting on an enterprise level Microsoft SQL Server platform.



**Figure 3.** Report form location (red dot with yellow 'C') with leaders connected to all boreholes associated with the report.





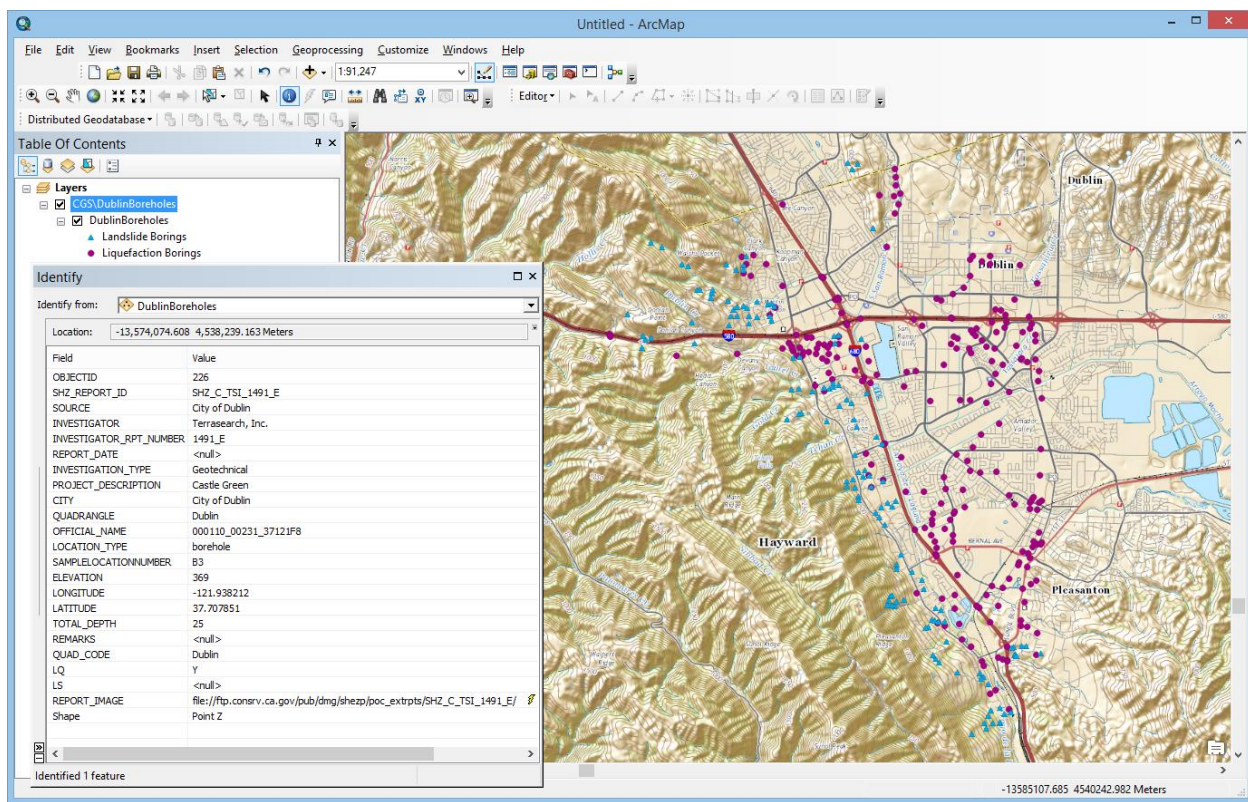
**Figure 4.** New geotechnical data model.

For all existing borehole data, new report locations had to be generated. For landslide data, CGS staff manually digitized new report locations, entered the report-level information, and created links to the scanned reports and Access boreholes.

For liquefaction data, data quality and completeness checks were done to identify gaps or conflicts in information required by the new data model. The legacy data were updated to correct the gaps/conflicts, then report locations and attributes were generated from the existing borehole database. The report information was then joined to an existing database of the scanned PDF file names and network locations to generate the paths needed for the PDF hyperlinks. The report locations, attribute information, and PDF links generated by the database joins were then manually verified by students overseen by CGS staff.

## Results:

Funding through the NGGDPP provided the impetus and opportunity to create data translation templates and scripts necessary to migrate the legacy borehole information into the new data model and link it to the scanned PDFs, thereby preserving the geotechnical data collected by CGS in a comprehensive enterprise database. In total, 1620 reports were created for the 2500 boreholes in Alameda County. Reports and boreholes for the Dublin 7.5-minute Quadrangle were selected to provide a proof of concept for web hosting (Figure 5).



**Figure 5. Feature service loaded into ArcMap for future public and staff access**

**Future:**

As a result of this project, CGS has a verified path for preservation of existing digital borehole data. The dataset tested in this project represents less than 5% of all the borehole data collected by CGS. As staff time becomes available, CGS will verify data preservation for other areas of the state.

Development of a final web service for CGS data is ongoing. CGS is currently working on a simple web based interface to display our current store of liquefaction borehole data. This interim interface will appear as a tab in our Information Warehouse <http://maps.conservation.ca.gov/cgs/informationwarehouse/>, and will remain live until development and deployment of the final ArcMap feature service.